CIRJE-F-1155

Vocational Training for Demobilized Ex-combatants with Disabilities in Rwanda

Kengo Igei
Graduate School of Economics, The University of Tokyo
Kana Takio
Harvard T. H. Chan School of Public Health
Keitaro Aoyagi
Metrics Work Consultants Inc.
Yoshito Takasaki
The University of Tokyo

September 2020

Vocational Training for Demobilized Ex-combatants with Disabilities in Rwanda*

Kengo Igei†, Kana Takio‡, Keitaro Aoyagi§, and Yoshito Takasaki**

August 8, 2020

Abstract

Disability-inclusive development is receiving growing attention as a pressing international development issue. Disability-inclusive development is especially urgent and complicated in post-conflict countries. This paper examines the impacts of vocational training on economic empowerment and social reintegration among demobilized ex-combatants with disabilities in Rwanda. This is the first quasi-experimental study on vocational training for disabled ex-combatants. Exploiting the variation in the timing of training uptake within the same training course, we employ a pipeline approach in the following three steps: (1) trimming to guarantee common support within courses, (2) exact matching on key covariates within courses, and (3) regression controlling for covariates within courses based on the matched sample. The results show that the training greatly increased not only employment and income, but also trainees' reintegration into the family and community. The results are robust to potential omitted variable bias and attrition bias according to a coefficient stability test and bound analysis, respectively. Our findings suggest a significant potential of vocational training for disabled ex-combatants in disarmament, demobilization, and reintegration programs. Our study exemplifies the utility of a credibly designed pipeline approach, which can be applied in a wide range of development projects in practice.

Keywords: Disability; Ex-combatants; Vocational training; Pipeline approach; Rwanda
JEL codes: I12, J24, O15, O17, Z13

---

* We wish to thank the Japan International Cooperation Agency (JICA) for their data on the project studied in this paper, "Skills Training for the Reintegration of Demobilized Soldiers with Disabilities," and Daisuke Sagiya for information about the project. We also wish to thank the Rwanda Demobilization and Reintegration Commission (RDRC) for their Rwanda Demobilization and Reintegration Programme (RDRP) data. Special thanks are owed to the Rwandan trainees who so willingly participated in the surveys. An earlier version of this paper benefited significantly from the comments and suggestions of Mayuko Onuki and participants at the 15th Annual Conference of the Japan Evaluation Society. This research has been made possible through support provided by the Japan Society for the Promotion of Science (25257106). Any errors of interpretation are solely the authors' responsibility. The first two authors share senior authorship.
† Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan. Email: igei@metricswork.co.jp
‡ Harvard T. H. Chan School of Public Health, 677 Huntington Ave, Boston, MA, 02115, USA. Email: kana.takio@gmail.com
§ Metrics Work Consultants Inc., 2-11-8 Yotsuya, Shinjuku-ku, Tokyo, 160-0004, Japan. Email: aoyagi@metricswork.co.jp
** Corresponding author. Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033 Japan. Phone: +81 3 5841 5530. Fax: +81 3 5841 5521. Email: takasaki@e.u-tokyo.ac.jp

1. Introduction

Disability-inclusive development has been receiving growing attention as a pressing international development issue (WHO and World Bank 2011; United Nations 2019). Confronted with physical, attitudinal, and institutional barriers against social participation and excluded from opportunities for education and employment, health care, and other basic services, persons with disabilities are often the poorest of the poor in societies (Banks, Kuper, and Polack 2017; Mitra 2018; Mitra, Posarac, and Vick 2013). Employment is increasingly recognized as a key factor for empowerment and inclusion in society. Vocational training for persons with disabilities has the potential to contribute to disability-inclusive development (ILO 2017). In particular, skill enhancement may strengthen their self-esteem and self-efficacy, thus reducing the self-stigmatization that could lead to low self-expectations about their employment prospects (Rohwerder 2018; WHO and World Bank 2011).

Disability-inclusive development is especially urgent and complicated in post-conflict countries. Following wars and conflicts, disabilities are likely to be prevalent among survivors, accompanied by social exclusion of persons with disabilities from basic services such as health care, food, water, and shelter (Handicap International 2015; Kett and van Ommeren 2009; Klyman, Kouppari, and Mukhier 2007). In post-conflict countries, disability is strongly associated with poverty across multiple dimensions (Mactaggart et al. 2019; Trani et al. 2011) and has long-lasting consequences for their livelihoods and well-being (Han, Dagsvik, and Cheng 2020; Palmer et al. 2019; Takasaki 2019). Disabled ex-combatants are confronted with social exclusion due to not only their disabilities, but also their ex-combatant status. Disarmament, demobilization, and reintegration (DDR) programs sponsored by United Nations agencies are major peace-building operations and vocational training is a major reintegration measure. Vocational training may have a significant potential to relieve the double burdens born by demobilized ex-combatants with disabilities through their economic empowerment and social reintegration.

Evidence for the effectiveness of vocational training for persons with disabilities in developing countries is scarce. Extant works mainly study youth, women, and unemployed people, but not those with disabilities; they also mainly focus on labor market outcomes, paying limited attention to social reintegration (Chinen et al. 2018; Escudero et al. 2019; McKenzie 2017; Tripney et al. 2013). A systematic review by Tripney et al. (2017) of labor market

programs for persons with physical or sensory disabilities in developing countries identified 14 studies, including four covering occupational rehabilitation, but none employed rigorous methods to remove bias due to the nonrandom assignment of treatment status. The authors concluded, 'The methodological inconsistencies and weaknesses of the current evidence base, and specific knowledge gaps, suggest a number of future research priorities' (Tripney et al. 2017, 7).

Evidence for the effectiveness of vocational training in post-conflict settings is also weak. Vocational training for ex-combatants in a Liberian DDR program was shown to increase employment, but not income (Levely 2014). Blattman and Ralston (2015) conclude that post-conflict vocational training has little impact on economic empowerment and social integration; however, their review covers no programs targeting persons with disabilities. This paper examines the impacts of vocational training on economic empowerment and social reintegration among demobilized ex-combatants with disabilities in Rwanda. To our knowledge, this is the first quasi-experimental study on vocational training for disabled ex-combatants.

We employ a pipeline approach, a quasi-experimental method also known as stepped-wedge design, phase-in design, or a waitlist study (e.g., Khandker, Koolwal, and Samad 2010; White and Raitzer 2017). Considering a situation in which some beneficiaries receive an intervention at a certain time and others will receive the same intervention later, the pipeline approach treats these two groups as treatment and comparison groups, respectively. This comparison between groups of beneficiaries is more credible than comparing beneficiaries and nonbeneficiaries. Examples of impact evaluation using the pipeline approach in developing countries include Coleman (1999) and subsequent studies on microfinance, Bose (2017) and Deininger and Liu (2019) on public employment, and Adoho et al. (2014) and Cho et al. (2013) on vocational training. Whereas none of these studies address disabilities, Grider and Wydick (2016) apply the pipeline approach to assess the impacts of providing a wheelchair to disabled people in Ethiopia.

We compared individuals who had just completed training (treatment group) with those who were about to start (comparison group). Although trainees in both groups passed through the same selection process, the assignment of the treatment status is determined by the timing of training uptake, which was not randomized, but not uptake per se. Following Imbens and Rubin (2015), to mitigate bias due to the nonrandom assignment of the timing of uptake, we employ the

following empirical design in three steps. In step 1, we focus on trainees who took the same course in two consecutive batches (i.e., common support within courses). In step 2, based on this trimmed sample, we conduct exact matching on key covariates, which are correlated with both the treatment status and outcomes, within courses to estimate average treatment effects on the treated. In step 3, based on the matched sample generated in step 2, we conduct a regression analysis controlling for additional covariates. The matching estimates in step 2 show that the training greatly increased not only employment and income, but also reintegration into the family and community; the regression results in step 3 are similar.

Our identifying assumption is that the timing of training uptake is independent of potential outcomes with covariates controlled for (i.e., unconfoundedness, or conditional independence). To address two potential threats to identification, we conduct two sensitivity analyses. First, omitted variable bias can negate conditional independence. We employ the coefficient stability test developed by Oster (2019), bounding the regression estimates in step 3 using a consistent estimator for the bias-adjusted treatment effect. The results suggest that the estimated treatment effects are robust to omitted variable bias.

Second, attrition systematically correlated with treatment status can cause bias. This attrition bias is an innate problem in the pipeline approach because of distinct survey attrition. In our setting, survey attrition is relevant for those in the treatment group who were tracked after the training (follow-up survey), but not for those in the comparison group who have not yet started the training (baseline survey). Previous studies employing the pipeline approach have not recognized or addressed such systematic attrition. We bound the matching estimates in step 2 by Lee bounds (Lee 2009). The two assumptions needed for Lee bounds are randomness of the treatment assignment and monotonicity—the treatment increases (or decreases) attrition for all individuals. Whereas we necessarily rely on conditional independence for the randomness in our quasi-experimental design, monotonicity is most likely to hold precisely because of the pipeline approach itself. This advantage of the pipeline approach has not been recognized in the literature. Although Lee's bound analysis has been commonly used to address attrition bias in both experimental and quasi-experimental studies, limited attention has been given to the validity of monotonicity in practice. The results show that the estimated treatment effects are robust to attrition bias.

2. Background and programs

2.1. Reintegration programs

After years of civil war between the Hutu and the Tutsi and conflicts with neighboring countries, the government of Rwanda has prioritized downsizing the Rwandan Patriotic Front (RPA), which was renamed the Rwanda Defence Forces (RDF) as the government army in 2002. In 1997 the Rwandan government established the Rwanda Demobilization and Reintegration Commission (RDRC) as an authority for planning and implementing the Rwanda Demobilization and Reintegration Programme (RDRP). With the support of international organizations, the RDRP was implemented in three stages: stage 1 (1997–2001) mainly targeted RPA ex-combatants; stage 2 (2002–2008) expanded the coverage to ex-combatants of the Forces Armées Rwandaises (the former government army) and private armed groups; and stage 3 (2009–2017) continued support for ex-combatants of private armed groups and further reduction of the RDF with particular attention to female, child, and disabled ex-combatants. Approximately 70,000 combatants were demobilized from 1997 to 2017, including about 9,000 who were disabled (RDRC 2017).

The RDRP initiated vocational training in 1998 and reinforced it over time (Finn, Baxter, and Onur 2014). Ex-combatants could receive a reintegration grant over six months after demobilization. Before receiving the grant, they had an opportunity to receive counseling and guidance to plan their reintegration and were required to explain how they would use their grant for services and necessities such as vocational training, income-generating activities, housing, and so on. They could receive additional support called the Vulnerability Support Window (VSW) if they met criteria such as having no shelter, land, assets, or skills. In 2007 the criteria were expanded to include permanent disability or chronic illness. The VSW was followed by screening for vocational training. In 2005, the RDRC conducted a survey among ex-combatants with disabilities who participated in the RDRP. It showed that 40% of respondents were employed, only 10% had a role in decision making within their families, and 70% felt they were treated as inferior in their communities (RDRC 2006).

2.2. JICA vocational training

As part of the RDRP, the Japan International Cooperation Agency (JICA) implemented a project in Rwanda called “Skills Training for the Reintegration of Demobilized Soldiers with Disabilities” (hereafter “the project”) from 2005 through 2008 (i.e., late stage 2). Our impact

evaluation focuses on this project. The project aimed at upholding Rwandan peace-building operations through vocational training for demobilized ex-combatants with disabilities. It also sought to promote their empowerment through social reintegration and recovery of self-esteem.[1] During the three-year period, a total of 27 types of training courses were offered at 10 training centers across the country, including in Kigali (the nation's capital) and other provinces. Depending on the capacity of the training centers, these courses covered a wide range of skills, such as welding, plumbing, carpentry, tailoring, cooking, and so forth. The duration typically ranged from six months to one year (except for a two-month agricultural course); the same courses were offered in sequence over three years.

The procedure for recruiting trainees was common across training centers and courses and similar to other vocational training programs in the RDRP described above. Upcoming courses were first advertised via radio, a common medium in Rwanda, and provincial and district RDRC reintegration officers cooperated in finding applicants for the training. The applicants went through an initial two-stage screening at training centers. In the first stage, demobilization ID, level of disability, and past training were checked. The applicants who passed the first-stage screening filled out interview and agreement forms, which were also used to assess their level of literacy, and chose a preferred training course. Then, they took a basic literacy and numeracy test prepared by project staff, followed by an interview to examine applicants' motivation for training and future work plans. Vocational training experts, directors and trainers of the training center, and officials of the provincial authorities of medical rehabilitation, reintegration, and monitoring and evaluation met to select trainees and assign training courses according to applicants' preference and disabilities. Most applicants who were accepted took up the training. Trainees who completed the program received a toolkit corresponding to their area of training to help them find a job with their acquired skills. In total, 925 persons with disabilities enrolled in the training and 867 graduated.

[1] The project targeted ex-combatants with disabilities regardless of their army group, formed according to ethnicity. With no data about social cohesion across ethnic groups, we cannot examine the training impacts on ethnic reintegration.

3. Empirical design

3.1. Pipeline approach

At each training center, each type of training course commonly consisted of multiple batches of the same course. A short survey was used to collect trainees' basic information; it was not designed for impact evaluation.[2] The timing of the surveys varied across training types and across batches within the same type: for some batches the survey was conducted before the training started and for other batches it was done after the training was completed. This variation allows us to employ the pipeline approach for evaluation because the surveys effectively serve as baseline and follow-up measures.

We focus on the baseline survey conducted for trainees at the timing of their enrollment in January and February 2008 and the follow-up survey conducted for graduates in June 2008. Figure 1 depicts the timeline of the training courses and surveys for batches in each of the nine training centers covered by either the baseline or follow-up survey (no survey was conducted at one of the centers). These surveys covered a total of 25 types of training courses in which 630 out of 925 trainees (68.1%) enrolled. These 630 enrolled trainees constitute the enrollment sample for our analysis.[3]

For each course, we consider graduated trainees covered by the follow-up survey as a treatment group and enrolled trainees covered by the baseline survey as a comparison group as depicted in Figure 1. The comparison group was enrolled at only four training centers (ATC, CFJ Kibali, CJF Nyanza, and CFJ Rwabuye). In this design, the assignment of the treatment status is not random because the timing of training uptake depends on individuals' decisions about the timing of their application, but not application per se. With about a six-month gap between the baseline and follow-up surveys, we can only capture the training impacts approximately six months after its completion; it is assumed that labor market conditions and other common factors affecting outcomes changed little over time. Below, our sensitivity analysis on omitted variable bias addresses an unobserved time trend.

---

[2] The project survey data were supplemented by the data of demobilized ex-combatants from the RDRC. As a result, we have almost complete data for the attributes of enrolled trainees used as covariates below.

[3] The remaining 295 enrolled trainees graduated before the baseline survey or after the follow-up survey, if they graduated. We exclude the former trainees from the enrollment sample because (1) we cannot construct their corresponding comparison group, or (2) even if we can do so for some of them, they are not in the batch consecutive to the comparison group and thus are trimmed from the sample as discussed below. The trainees who graduated after the follow-up survey were not covered by the baseline survey.

The surveys covered 64.6% (407/630) of enrolled trainees. The proportion of surveyed trainees is quite different between the comparison and treatment groups: 93.2% (150/161) versus 54.8% (257/469), meaning that the attrition rate in the treatment group was 38.4% higher than in the comparison group (Table 1, column 1). This is mostly because the follow-up survey failed to track a considerable number of graduated trainees due to, for example, migration.[4] We confirm that attrition is significantly higher in the treatment group than the comparison group with the training center-course (strata) fixed effects and other covariates (used in the main analysis discussed below) controlled for; attrition is also significantly correlated with the center-course fixed effects and covariates (Table 1, column 2). Below, we conduct a sensitivity analysis to address potential attrition bias.

The covariate balance between the treatment and comparison groups among surveyed trainees is shown in Table 2, where the difference in means is tested based on robust standard errors.[5] Physical disability excluding amputation (55.5%), amputation (12.0%), and blindness (8.8%) are common.[6] Although age, marital status, and years since demobilization are well balanced, education level, type of disability, and ex-army groups are not balanced (columns 1–3). The results for enrolled trainees (not shown) are very similar, indicating that this imbalance is not caused by survey attrition.

To mitigate bias due to the nonrandom assignment of treatment status, we employ the following three-step empirical design suggested by Imbens and Rubin (2015). In step 1, we focus on surveyed trainees who took the same training course in two consecutive batches by trimming the surveyed sample through stratification by course at the training center. This guarantees the common support of propensity scores, the probability of being treated, within center-course strata. The trimmed sample includes 94 and 114 trainees in the treatment and comparison groups, respectively, covering a total of 13 types of courses at the four training centers mentioned above (Figure 1 and Appendix, Table A1). The covariate balance in the trimmed sample (Table 2, columns 4–6) is similar to the original results.

In step 2, based on this trimmed sample, we employ a matching approach with the

[4] Other reasons for attrition in the treatment group include dropout and death. Attrition in the comparison group was only due to refusal of the survey.

[5] The results with center-course fixed effects controlled for in regression (not reported) are very similar.

[6] Other disabilities (17.0%) include retained foreign bodies, mental illness, and multisystemic lesions. With multiple disabilities, the dummy variables for each type of disability are not mutually exclusive.

conditional independence assumption that the treatment status (the timing of training uptake) is independent of potential outcomes with covariates controlled for. Since all unbalanced covariates—education level, type of disability, and ex-army group—are discrete variables, we employ exact matching using these covariates within center-course strata.[7] Within-strata matching is crucial because the selection of trainees and the assignment of training courses were done at each training center. In other words, we assume that the exact matching on these covariates mimic blocked (stratified) randomization (King and Nielsen 2019). The matched sample includes 43 and 38 trainees in the treatment and comparison groups, respectively, covering a total of 11 types of courses at the four training centers (Figure 1 and Appendix, Table A1). Given the conditional independence assumption, an unbiased estimator for average treatment effect on the treated is the difference in weighted means between the treatment and comparison groups in the matched sample, using strata size as weights,[8] which is given by the estimated coefficient of the treatment dummy in the weighted univariate regression (Gerber and Green 2012). Here, all we can identify is the impacts of the mix of the 11 types of courses in the matched sample, the composition of which is somewhat different from that of the 27 types in the project.[9] Our analysis does not say anything about the impacts of courses other than these 11 types remaining in the matched sample, and thus the effectiveness of the project as a whole. This limitation is unavoidable in our empirical design for internal validity.

With the small number of observations in the matched sample, weak statistical power is a concern. At the same time, it is important to consider clustering in inference. In addition to robust standard errors, we use two-way cluster-robust standard errors to capture error terms correlated within centers and courses. When the number of clusters is small, the cluster-robust variance estimator overrejects null hypotheses. With a small number of centers (4) and courses (11), we employ a wild-cluster bootstrap (which is reliable for a small number of clusters; Cameron, Gelbach, and Miller 2008) of the two-way cluster-robust standard errors (Roodman et

---

[7] For education level and type of disability, we consider missingness as a separate category.

[8] The weight for the treatment group is equal to one, and the weight for the comparison group is equal to $m_C/m_T \cdot m_T^s/m_C^s$, where m_T and m_C are the total number of matched observations of the treatment and comparison groups, respectively, and m_T^s and m_C^s are the number of matched observations of the treatment and comparison groups, respectively, in stratum s.

[9] These 11 types of courses consist of brick/block laying and concreting at three centers; welding at two centers; plumbing and pipe fitting at two centers; and electronics, cooking, carpentry and joinery, and tailoring at one center each (Appendix, Table A1). The remaining 16 types include courses other than these 11 types, such as computing, shoemaking, and leather craft.

al. 2019).[10]

In step 3, based on the matched sample, we conduct a regression analysis controlling for additional covariates—age, marital status, years since demobilization—as well as the matching variables, including center-course strata fixed effects.[11] This regression analysis in the matched sample is more credible than the naïve (conventional) regression analysis based on the surveyed sample (before step 1) or the trimmed sample (after step 1).

3.2. Outcomes

Our outcomes of interest are economic empowerment and social reintegration, both of which are measured by respondents' self-report in the survey. Economic empowerment is captured by employment and income. We construct a dummy variable for employment, which takes 1 if the employment status is "self-employed in the nonagriculture sector," "employed by employers," or "a cooperative member," and 0 if the status is "self-employed in the agriculture sector," "unemployed," "retired," or "unable to work." Income is measured by the average monthly income from all sources except government grants. Respondents with the employment dummy taking 0 can also earn income because they can have earnings in the agriculture sector, petty trading, and other minor activities and receive remittances from family and others. We use both the income level and its natural logarithm, the latter of which is more robust to outliers.

Social reintegration is captured by respondents' feelings about their relationships with their family, neighbors, and people in the community. First, feelings toward family were measured by the question, "Since demobilization, do you feel you are treated as inferior by your family?" with three options: "treated as inferior," "treated the same," or "treated preferably." We use a dummy variable that takes 1 if the answer is "treated preferably" and 0 otherwise (the results for an alternative dummy variable that takes 1 if the answer is "treated preferably" or "treated the same" and 0 otherwise are qualitatively the same). Second, feelings toward neighbors were measured by the question, "How would you describe your relationships with your neighbors?" with four options: "good," "average," "bad," or "not sure." We use a dummy

[10] When computing standard errors, we adopt Webb's (2014) six-point distribution as the auxiliary distribution and cluster bootstrap errors by training center, which is a cluster with the smallest number, following MacKinnon, Nielsen, and Webb (2020). We do 1000 bootstrap replications.

[11] Marital status includes a category for missingness. We combine Forces Armées Rwandaises and armed group into one group due to the small number of observations of the latter. Exact matching based on this dummy variable yields the same matching results.

variable that takes 1 if the answer is "good" and 0 otherwise (the "bad" category was uncommon and the "not sure" category was dropped). Third, feelings toward people in the community were measured by the question, "Since demobilization, do you feel you are treated as inferior than other men and women in your community?" with three options: "treated as inferior," "treated the same," or "treated preferably." We use a dummy variable that takes 1 if the answer is "treated the same" or "treated preferably" and 0 otherwise (the "treated preferably" category was uncommon). A potential concern about these subjective questions is social desirability bias: respondents might have answered in a way that would make them be seen as favored by others in the society, the surveyors, or the researchers. In our context, trainees in the treatment group might have pretended that they were accepted by the family or in the community after the training. Below, our sensitivity analysis on omitted variable bias addresses these potential measurement errors.

Among surveyed trainees, missing values are common for income (28.7%) and relationship with the family (14.5%), but not for the other outcomes. This causes additional attrition. Opposite to the survey attrition discussed above, the nonresponse rates of these two outcomes are lower in the treatment group than in the comparison group; qualitatively the same comparison holds in the trimmed and matched samples (Appendix, Table A2). We return to this systematic attrition due to missingness when we interpret estimation results.

Thus, we consider three economic empowerment outcomes and three social reintegration outcomes. When multiple hypothesis tests are conducted, a major concern is overrejection: The probability of falsely rejecting a true null hypothesis—the probability of a type I error—exceeds the significance level. We employ multiple hypothesis correction, proposed by Romano and Wolf (2005a, 2015b), to control the familywise error rate, the probability of falsely rejecting at least one true null hypothesis among a set of hypotheses. This correction has an advantage over the conventional Bonferroni or Holm correction in its ability to correctly reject a false null hypothesis because it allows the dependence among the p-values or the test statistics from multiple hypothesis testing with resampling and step-down procedure. We adjust p-values for each estimate of the treatment effect within each category of economic empowerment and social reintegration outcomes. The correction of p-values is based on the robust standard errors with 1,000 bootstrap replications.

4. Results

4.1. Main results

Table 3 reports the means of outcomes in the comparison group based on the matched sample (those in the trimmed sample are similar), the matching estimates of the treatment effects in step 2 (panel A), and the regression estimates of the treatment effects in step 3 (panel B). In the comparison group, 18.5% were employed and the mean monthly income was about 8,000 Rwandan francs, which is equivalent to about 14 US dollars (1US$ equaled 547 RWF in 2008). This indicates that before the training, ex-combatants with disabilities were living in severe poverty. The extent of reintegration into the family and community before the training was weak. These patterns are consistent with those found in two previous studies: the RDRC (2006) study discussed above and a World Bank survey for ex-combatants who completed RDRP vocational training and those who did not participate in it (Finn, Baxter, and Onur 2014).[12]

The matching estimates of the treatment effects on all outcomes but relationships with neighbors are positive, with statistical significance at conventional levels based on robust standard errors (reported in parentheses; panel A). These statistical results are robust to two-way cluster-robust standard errors (p-values in brackets) and Romano-Wolf correction for multiple hypothesis testing (p-values in braces). The statistical results of the regression estimates are similar to the matching estimates (panel B). Although the estimate for income in level loses statistical significance based on robust standard errors, it is statistically significant based on two-way cluster-robust standard errors. Almost all matching estimates with statistical significance are similar to the regression estimates. As an exception, the latter estimate for income in level is somewhat smaller than the former estimate, though the estimates for log income are very similar. The results for income in level might be due to outliers. Overall, these results suggest that if the conditional independence holds in the regression (step 3), it is also likely to hold in the matching (step 2).

The training increased employment by over 0.6 in probability (almost 3.5 times the

[12] According to Finn, Baxter, and Onur (2014), ex-combatants who completed vocational training were more likely to have work experience after demobilization and income from nonagricultural activities than those who did not participate in vocational training. They also had more confidence about themselves and felt more trusted by their family and community. In their sample, 25% of ex-combatants who completed vocational training were disabled and their labor market outcomes did not differ significantly by disability status. The purpose of their study was not to evaluate the impacts of RDRP vocational training.

comparison mean) and income by about 90%. The training increased the probability of being treated preferably by family and treated the same or preferably in the community by over 0.3 (about 1.5 and 1.8 times, respectively, the comparison mean). The training did not significantly increase the probability of having good relationships with neighbors, probably because the incidence of good relationships was relatively high before the training (comparison mean: 75%). Despite the weak power with the small sample size, these estimates are statistically significant because the effect sizes are sufficiently large. These results are consistent with the comparison of the distributions of income (log) and relationships with family, with neighbors, and in the community between the treatment and comparison groups in the matched sample depicted in Figure 2. Hence, the vocational training was effective in improving not only employment and income, but also reintegration into the family and community.

We speculate how the systematic attrition due to missingness of income and relationships with the family affects the treatment effect estimates. Suppose that respondents' economic conditions and family relationships were more unsatisfactory in the comparison group than in the treatment group partly due to positive treatment effects, which we consider likely. Then, those in the comparison group might have hesitated more to respond to the corresponding survey questions. In this case, systematic attrition due to missingness biases the estimated treatment effects downward and these positive estimates are qualitatively robust to such bias. These results are also qualitatively the same as those for related outcomes: employment and relationships in the community with limited missing values.

When we repeat the analysis in the trimmed sample before matching (step 1), all the estimated treatment effects, including those for income in level and relationship with neighbors, are statistically significant; in particular, the estimate for relationships with neighbors is about two times the original result (Appendix, Table A3). This contrast indicates the importance of balancing the treatment and comparison groups through exact matching for identification.

4.2. Coefficient stability test

Omitted variable bias, including an unobserved time trend within center-course strata and systematic measurement errors such as social desirability bias, can negate the conditional independence assumption for identification. We employ the coefficient stability test against omitted variable bias in the regression framework developed by Oster (2019). The test is based

on the movements not only in the key coefficient, but also in R-squared values when additional covariates are included. The bias-adjusted treatment effect is approximately given by:

$$\beta^* \approx \tilde{\beta} - \delta(\dot{\beta} - \tilde{\beta}) \frac{R_{max} - \tilde{R}}{\tilde{R} - \dot{R}}$$

where $\tilde{\beta}$ and $\tilde{R}$ are the estimate of treatment effect and R-squared value from the regression with observable covariates (step 3); $\dot{\beta}$ and $\dot{R}$ are the estimate of treatment effect and R-squared value from the regression with no covariates (step 2); and R_{max} is the R-squared value from the regression with a complete set of covariates, including unobservable ones, that is hypothetical and R_{max} has to be set by analysts. δ is the degree of selection on unobservables relative to observables, resulting from the assumption of the proportional relationship between selection on unobservables and observables, $\delta \frac{Cov(Treated,Unobservables)}{Var(Unobservables)} = \frac{Cov(Treated,Observables)}{Var(Observables)}$. Following the procedure proposed by Oster (2019), we compute the bias-adjusted treatment effect on each outcome for the following four cases: (R_{max}, δ) = (1.3$\tilde{R}$, 1), (1.3$\tilde{R}$, -1), (2$\tilde{R}$, 1), and (2$\tilde{R}$, -1).[13] If the bounding set $[\tilde{\beta}, \beta^*]$ includes zero or β^* is outside the confidence interval of $\tilde{\beta}$, we can conclude that the estimated treatment effect in the regression analysis (step 3) is not robust to the omission of unobservables.

Table 4 presents the estimates and R-squared values from the regressions with no covariates ($\dot{\beta}$, $\dot{R}$) and a full set of covariates ($\tilde{\beta}$, $\tilde{R}$) shown in Table 3, the confidence intervals obtained by robust standard errors, and the bias-adjusted treatment effects (β^*) based on each case of (R_{max}, δ). We omit income in level and relationships with neighbors, the estimated treatment effects of which were found to be not consistently significant above. The bounding set $[\tilde{\beta}, \beta^*]$ does not include zero for any outcomes in any case of (R_{max}, δ), and β^* is inside the confidence interval of $\tilde{\beta}$ for all outcomes in all cases of (R_{max}, δ). These results suggest that the regression estimates of treatment effects are qualitatively robust to omitted variable bias, supporting the conditional independence in the regression estimates and thus in the matching estimates.

[13] Oster (2019) suggests δ=1 as an upper bound, which indicates that the degree of selection on unobservables is no greater than that on observables. In our setting, this assumption seems to be appropriate because the exact matching should limit the difference in unobservables between the treatment and comparison groups. We also consider δ=-1, corresponding to the case where the direction of selection on unobservables is opposite to that on observables.

4.3. Lee bounds

In our pipeline approach, survey attrition is necessarily more likely to happen in the treatment group than the comparison group, as discussed above. This is confirmed in the surveyed sample and the trimmed and matched samples (this is the case with the center-course strata fixed effects and covariates controlled for; Table 1). To deal with potential bias caused by this systematic attrition, we estimate Lee bounds for the matching estimates in step 2 (Lee 2009). Monotonicity is likely to hold precisely because of this systematic attrition due to survey attrition. We assume conditional independence in the matching estimates; then, the randomness of the assignment of treatment status holds. As in the coefficient stability test, we omit income in level and relationships with neighbors. We do not include observations with missing outcomes, which are common for income in log and relationships with the family, because monotonicity is unlikely to hold for attrition due to missingness: contrary to survey attrition, missingness was less common in the treatment group than the comparison group, as discussed above (Appendix, Table A2). Thus, our bound analysis addresses potential attrition bias solely due to survey attrition.

We first conduct within-strata exact matching in the same way as above among surveyed trainees in the trimmed sample (step 1) plus survey attritors within the same strata as those of the surveyed trainees. Based on the resulting matched sample including attritors, we compute the amount of trimmed tail of the distribution of outcome variable in each center-course stratum. That is, we identify matched attritors within strata to compute trimming proportion in a way that is consistent with the original analysis (Table 3, panel A).[14] We consider best-case and worst-case scenarios. The best-case scenario assumes that those who were surveyed because they were in the comparison group lie in the upper tail of the distribution of outcome variable. We trim the upper tail and compare the mean outcome of the remaining trainees of the comparison group with that of the treatment group to obtain the upper bound of the treatment effect. In the case of the worst-case scenario, we trim the lower tail of the distribution for the comparison group to obtain the lower bound of the treatment effect. In the matched sample, we apply this trimming procedure within each stratum for the trimming proportions obtained above. To be consistent with the original analysis (Table 3, panel A), we use the strata size in the original analysis as weights.

[14] The trimming proportion is defined as $q = (q_C - q_T)/q_C$, where q_C and q_T are the rates of nonattrited trainees in the comparison and treatment groups, respectively.

The estimates of Lee's lower and upper bounds for the treatment effects on the four outcomes are reported in Table 5.[15] For all outcomes, the lower bounds of the treatment effects are positive and significantly different from zero.[16] These results suggest that the matching estimates of treatment effects are qualitatively robust to attrition bias.

5. Discussion and conclusions

This paper evaluated vocational training for demobilized ex-combatants with disabilities in Rwanda. We found that the training improved not only employment and income, but also reintegration into the family and community about six months after the completion of the training. The results are robust to potential omitted variable bias and attrition bias. A caveat is that the results are based on selected courses rather than the project as a whole.

These results are distinct from weak impacts of vocational training for nondisabled people commonly found in extant studies (McKenzie 2017; Chinen et al. 2018). This difference can be attributed partly to the severity of disadvantages faced by persons with disabilities. Before the training, most ex-combatants with disabilities were unemployed with low income and felt that they were not treated well by family members and in the community. That is, their starting point was so low that the returns to vocational training could be large. This suggests that the effectiveness of vocational training depends on labor market conditions and social contexts.[17] At the same time, our evaluation of trainees does not capture nontrainees. Application for training was voluntary and careful screening of applicants was done based on literacy, numeracy, motivation, and future plans; that is, work readiness. The positive selection of trainees through self-selection and screening could underlie the training's strong impacts. This raises a potential concern about the generalizability of our findings for the population of ex-combatants with disabilities in Rwanda, though addressing external validity is beyond the scope of this paper.

Although disabled ex-combatants are confronted with the double burdens of being an ex-combatant and disabled, they have often been excluded from DDR programs, including

---

[15] The numbers of observations for income in log and the dummy for being treated preferably by family are slightly smaller than those in the original analysis because one or a few strata that lose common support due to missing values in these outcomes were dropped. For all outcomes, all attritors are in the treatment group.

[16] The systematic attrition due to missingness for income in log and relationships with the family might make the bound estimates biased downward, as discussed for the treatment effect estimates above. If this is the case, the corresponding lower bound estimates are likely to be qualitatively robust to such bias.

[17] Kluve et al. (2019) find larger effects of active labor market programs in low- and middle-income countries than high-income countries, discussing a possibility of lower starting points in the former as a reason.

vocational training (Handicap International 2015; ILO 2009; Lord and Stein 2015). This exclusion might be underlain by the biomedical framework of international humanitarian actions, which considers disability a heath issue and focuses on medical care and rehabilitation (Lord and Stein 2015). Program planners also tend to overlook the desire of persons with disabilities to work and acquire skills to be productive (ILO 2009). Our findings shed new light on the significant potential of vocational training to improve economic empowerment and social reintegration among disabled ex-combatants in DDR programs toward disability-inclusive development in post-conflict countries.

Our pipeline approach exploited the variation in the timing of training uptake across consecutive batches within the same course. Our empirical design consisted of three steps: (1) trimming to guarantee common support within courses, (2) matching on key covariates within courses, and (3) regression controlling for covariates within courses based on the matched sample. To address omitted variable bias and attrition bias, we conducted two sensitivity analyses: a coefficient stability test (Oster 2019) and Lee bounds (Lee 2009). Lee's bound analysis is effective to address attrition bias in the pipeline approach, for which systematic attrition is an innate problem. By executing the recommended practices of the pipeline approach, our study exemplified its utility as a quasi-experimental method, which in our view has been undervalued in the literature. The pipeline approach can be widely applied because it does not need to set a comparison group in advance and it can be used with data from beneficiary surveys, including those not designed for impact evaluation like ours, commonly used in development projects. These advantages are especially significant in cases where stakeholders become interested in its impacts after the project has started, or an experimental approach is not feasible for some reason. Impact evaluations based on a credibly designed pipeline approach can effectively inform policymakers on a wide range of development projects.

References

Adoho, F., S. Chakravarty, D. T. Korkoyah, Jr., M. Lundberg, and A. Tasneem. 2014. "The Impact of an Adolescent Girls Employment Program: The EPAG project in Liberia." World Bank Policy Research Working Paper 6832, World Bank, Washington DC. Accessed 7 August 2020. http://documents.worldbank.org/curated/en/610391468299085610/The-impact-of-an-adolescent-girls-employment-program-the-EPAG-project-in-Liberia.

Banks, L. M., H. Kuper, and S. Polack. 2017. "Poverty and Disability in Low- and Middle-Income Countries: A Systematic Review." *PLOS ONE* 12(12): e0189996. doi: https://doi.org/10.1371/journal.pone.0189996.

Blattman, C., and L. Ralston. 2015. “Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs.” Social Science Research Network. Accessed 7 August 2020. https://ssrn.com/abstract=2622220.

Bose, N. 2017. “Raising Consumption through India’s National Rural Employment Guarantee Scheme.” *World Development* 96: 245–263. doi: https://doi.org/10.1016/j.worlddev.2017.03.010.

Cameron A. C., J. B. Gelbach, and D. L. Miller. 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *Review of Economics and Statistics* 90 (3): 414–427. doi: https://doi.org/10.1162/rest.90.3.414.

Chinen, M., T. De Hoop, M. Balarin, L. Alcázar, J. Sennett, and J. Mezarina. 2018. “Vocational and Business Training to Improve Women’s Labour Market Outcomes in Low-and Middle-Income Countries: A Systematic Review.” 3ie Systematic Review 40, International Initiative for Impact Evaluation (3ie), London. doi: https://doi.org/10.23846/SR71094.

Cho, Y., D. Kalomba, A. M. Mobarak, and V. Orozco. 2013. “Gender Differences in the Effects of Vocational Training: Constraints on Women and Dropout Behavior.” World Bank Policy Research Working Paper 6545, World Bank, Washington DC. Accessed 7 August 2020. http://documents.worldbank.org/curated/en/882971468272376091/Gender-differences-in-the-effects-of-vocational-training-constraints-on-women-and-drop-out-behavior.

Coleman, B. 1999. “The Impact of Group Lending in Northeast Thailand.” *Journal of Development Economics* 60(1): 105–141. doi: https://doi.org/10.1016/S0304-3878(99)00038-3.

Deininger, K., and Y. Liu. 2019. “Heterogeneous Welfare Impacts of National Rural Employment Guarantee Scheme: Evidence from Andhra Pradesh, India.” *World Development* 117: 98–111. doi: https://doi.org/10.1016/j.worlddev.2018.12.014.

Escudero, V., J. Kluve, E. L. Mourelo, and C. Pignatti. 2019. “Active Labour Market Programmes in Latin America and the Caribbean: Evidence from a Meta-Analysis.” *The Journal of Development Studies* 55(12): 2644–2661. doi: https://doi.org/10.1080/00220388.2018.1546843.

Finn, A., D. Baxter, and M. Onur. 2014. *Making Vocational Training Work: A Study of Vocational Training in DDR*. Washington DC: The World Bank. Accessed 7 August 2020. http://documents.worldbank.org/curated/en/447321467991969774/Rwanda-Making-vocational-training-work-a-study-of-vocational-training-in-DDR.

Gerber, A. S., and D. P. Green. 2012. *Field experiments: Design, analysis, and interpretation*. New York: Norton.

Grider, J., and B. Wydick. 2016. “Wheels of Fortune: The Economic Impacts of Wheelchair Provision in Ethiopia.” *Journal of Development Effectiveness* 8(1): 44–66. doi: https://doi.org/10.1080/19439342.2015.1064986.

Han, X., J. Dagsvik, and Y. Cheng. 2020. “Disability and Job Constraint in Post Civil War Cambodia.” *The Journal of Development Studies*. Advance online publication. doi: https://doi.org/10.1080/00220388.2020.1769073.

Handicap International. 2015. *Disability in Humanitarian Context: Views from Affected People and Field Organisations*, London: Handicap International. Accessed 7 August 2020. https://reliefweb.int/report/world/disability-humanitarian-context-views-affected-people-and-field-organisations.

Imbens, G. W., and D. B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.

ILO (International Labor Organization). 2009. *Socio-economic Reintegration of Ex-combatants: Guidelines*. Geneva: ILO. Accessed 7 August 2020. https://www.ilo.org/wcmsp5/groups/public/@ed_emp/documents/instructionalmaterial/wcms_141276.pdf.

ILO (International Labor Organization). 2017. *Policy Brief: Making TVET and Skills Systems Inclusive of Persons with Disabilities*. Geneva: ILO. Accessed 7 August 2020. https://www.ilo.org/wcmsp5/groups/public/---ed_emp/---ifp_skills/documents/publication/wcms_605087.pdf.

Kett, M., and M. van Ommeren. 2009. "Disability, Conflict, and Emergencies." *LANCET* 374(9704): 1801–1803. doi: https://doi.org/10.1016/s0140-6736(09)62024-9.

Khandker, S., G. B. Koolwal, and H. A. Samad. 2010. *Handbook on Impact Evaluation: Quantitative Methods and Practices*, Washington, DC: World Bank.

King, G., and R. Nielsen. 2019. "Why Propensity Scores Should not be Used for Matching." *Political Analysis* 27(4): 435–454. doi: https://doi.org/10.1017/pan.2019.11.

Kluve, J., S. Puerto, D. Robalino, J. M. Romero, F. Rother, J. Stöterau, F. Weidenkaff, and M. Witte. 2019. "Do Youth Employment Programs Improve Labor Market Outcomes? A Quantitative Review." *World Development* 114: 237–253. doi: https://doi.org/10.1016/j.worlddev.2018.10.004.

Klyman, Y., N. Kouppari, and M. Mukhier. 2007. "Disability and Disasters towards an Inclusive Approach." Chap. 4 in *World Disasters Report 2007: Focus on discrimination*. Geneva: International Federation of Red Cross and Red Crescent Societies. Accessed 7 August 2020. https://www.ifrc.org/PageFiles/99876/2007/WDR2007-English.pdf.

Lee, D. S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76: 1071–1102. doi: https://doi.org/10.1111/j.1467-937X.2009.00536.x.

Levely, I. 2014. "Measuring Intermediate Outcomes of Liberia's Disarmament, Demobilization, Rehabilitation and Reintegration program." *Defence and Peace Economics* 25(2): 139–162. doi: https://doi.org/10.1080/10242694.2012.727065.

Lord, J. E., and M. A. Stein. 2015. "Peacebuilding and Reintegrating Ex-combatants with Disabilities." *The International Journal of Human Rights* 19(3): 277–292. doi: https://doi.org/10.1080/13642987.2015.1031515.

MacKinnon, J. G., M. Ø. Nielsen, and M. D. Webb. 2020. "Wild Bootstrap and Asymptotic Inference with Multiway Clustering." *Journal of Business and Economic Statistics.* Advance online publication. doi: https://doi.org/10.1080/07350015.2019.1677473.

Mactaggart, I., N. S. Maung, C. T. Khaing, H. Kuper, and K. Blanchet. 2019. "A Case-control Study of Musculoskeletal Impairment: Association with Socio-economic Status, Time Use and Quality of Life in Post-conflict Myanmar." *BMC Public Health* 19: 1502. doi: https://doi.org/10.1186/s12889-019-7851-5.

McKenzie, D. 2017. "How Effective are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence." *The World Bank Research Observer* 32(2): 127–154. doi: https://doi.org/10.1093/wbro/lkx001.

Mitra, S. 2018. *Disability, health and human development*. New York: Palgrave MacMillan.

Mitra, S., A. Posarac, and B. Vick. 2013. "Disability and Poverty in Developing Countries: A Multidimensional Study." *World Development* 41(1): 1–18. doi: http://dx.doi.org/10.1016/j.worlddev.2012.05.024

Oster, E. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence."

Journal of Business and Economic Statistics 37(2): 187–204. doi: https://doi.org/10.1080/07350015.2016.1227711.
Palmer, M., C. V. Nguyen, S. Mitra, D. Mont, and N. E. Groce. 2019. "Long-lasting Consequences of War on Disability." *Journal of Peace Research* 56(6): 860–875. doi: https://doi.org/10.1177/0022343319846545.
Rohwerder, B. 2018. "Disability Stigma in Developing Countries." K4D Helpdesk Report, Institute of Development Studies, Brighton. Accessed 7 August 2020. https://assets.publishing.service.gov.uk/media/5b18fe3240f0b634aec30791/Disability_stigma_in_developing_countries.pdf.
Romano, J. P., and M. Wolf. 2005a. "Exact and Approximate Stepdown Methods for Multiple Hypothesis Testing." Journal of the American Statistical Association 100(469): 94–108. doi: https://doi.org/10.1198/016214504000000539.
Romano, J. P., and M. Wolf. 2005b. "Stepwise Multiple Testing as Formalized Data Snooping." *Econometrica* 73(4): 1237–1282. doi: https://doi.org/10.1111/j.1468-0262.2005.00615.x.
Roodman, D., J. G. MacKinnon, M. Ø. Nielsen, and M. D. Webb. 2019. "Fast and Wild: Bootstrap Inference in Stata Using Boottest," *Stata Journal* 19: 4–60. doi: https://doi.org/10.1177/1536867X19830877.
RDRC (Rwanda Demobilization and Reintegration Commission). 2006. *Annual Report for 2005*. Accessed 7 August 2020. http://tdrp.net/mdrp/PDFs/Sum_RDRP_05_Final_rep.pdf.
RDRC (Rwanda Demobilization and Reintegration Commission). 2017. *Second Emergency Demobilization and Reintegration Project (SEDRP-P112712) Final Evaluation: Independent External Evaluation Final Report 2017*. Accessed 7 August 2020. http://demobrwanda.gov.rw/fileadmin/templates/demoba/documents/2006/Reports/Final_External_Independent_Evaluation_Report_2017.pdf.
Takasaki, Y. 2019. "Disability and Poverty: Landmine Amputees in Cambodia." CIRJE Discussion Paper CIRJE-F-1118, Center for International Research on the Japanese Economy, Tokyo. Accessed 7 August 2020. http://www.cirje.e.u-tokyo.ac.jp/research/dp/2019/2019cf1118.pdf.
Trani, J. F., J. Browne, M. Kett, O. Bah, T. Morlai, N. Bailey, and N. Groce. 2011. "Access to Health Care, Reproductive Health and Disability: A Large Scale Survey in Sierra Leone." *Social Science and Medicine* 73(10): 1477–1489. doi: http://dx.doi.org/10.1016/j.socscimed.2011.08.040.
Tripney, J., J. Hombrados, M. Newman, K. Hovish, C. Brown, K. Steinka-Fry, and E. Wilkey. 2013. "Technical and Vocational Education and Training (TVET) Interventions to Improve the Employability and Employment of Young People in Low-and Middle-income Countries: A Systematic Review." *Campbell Systematic Reviews* 2013:9. doi: https://doi.org/10.4073/csr.2013.9.
Tripney, J., A. Roulstone, C. Vigurs, N. Hogrebe, E. Schmidt, and R. Stewart. 2017. "Interventions to Improve the Labour Market for Adults Living with Physical and/or Sensory Disabilities in Low-and M-income Countries: A Systematic Review." 3ie Systematic Review 39, International Initiative for Impact Evaluation (3ie), London. Accessed 7 August 2020. http://3ieimpact.org/sites/default/files/2019-01/sr39-labour-market_0.pdf.
United Nations. 2019. *Disability and Development Report: Realizing the Sustainable Development Goals by, for and with Persons with Disabilities*, New York: United Nations. Accessed 7 August 2020. https://social.un.org/publications/UN-Flagship-Report-Disability-Final.pdf.

Webb, M. D. 2014. “Reworking Wild Bootstrap Based Inference for Clustered Errors.” QED Working Paper 1315, Queen’s University, Department of Economics, Ontario. Accessed 7 August 2020. http://qed.econ.queensu.ca/working_papers/papers/qed_wp_1315.pdf.
White, H., and Raitzer, D. A. 2017. *Impact Evaluations of Development Interventions: A Practical Guide*. Manila: Asian Development Bank.
WHO (World Health Organization) and World Bank. 2011. *World Report on Disability*, Washington DC: World Health Organization.
World Bank. 2013. *World Development Report 2013: Jobs*. Washington DC: World Bank.

Figure 1. Timeline

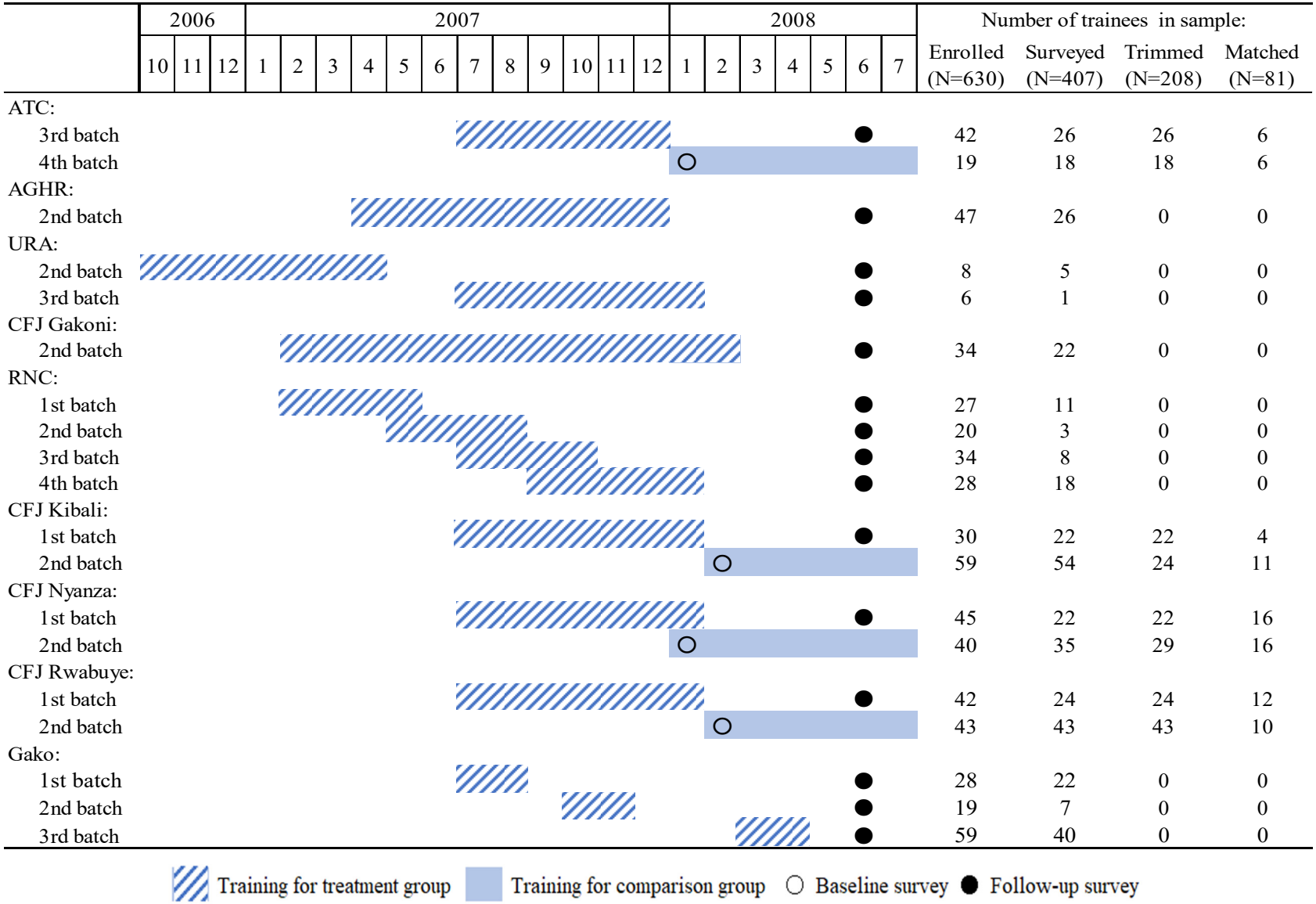
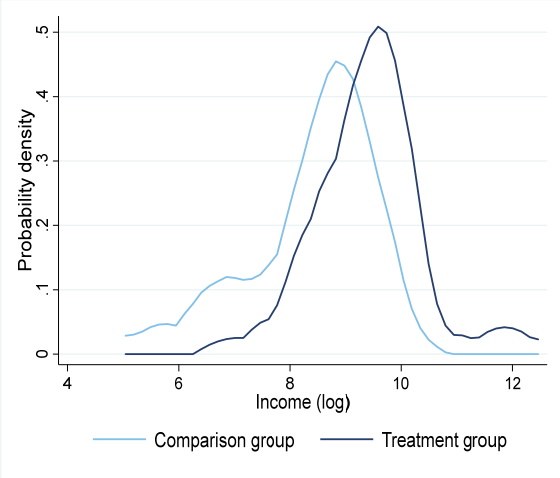
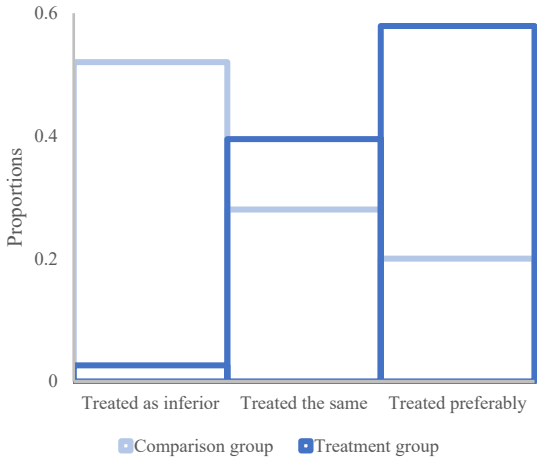


Figure 2. Distribution of outcome variables

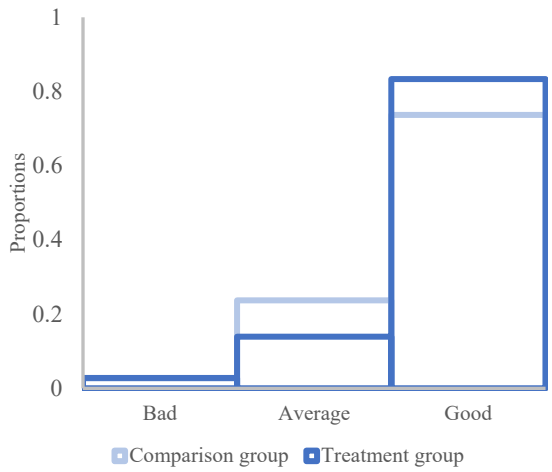
Panel A. Income (log)



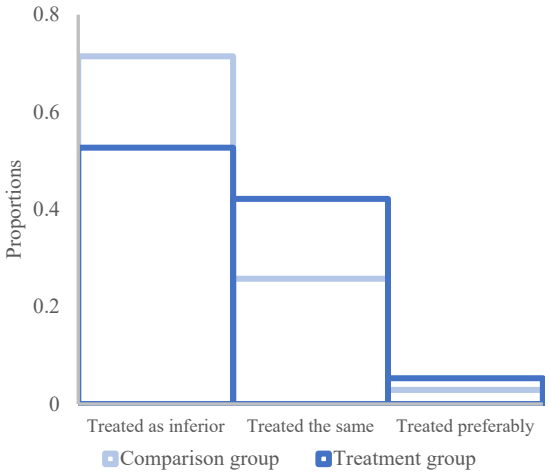
Panel B. Treatment by family



Panel C. Relationships with neighbors



Panel D. Treatment in community



Note: The sample is the matched sample.

Table 1. Attrition

	Enrollment sample		Trimmed sample		Matched sample	
	No covariates	With strata FE & covariates	No covariates	With strata FE & covariates	No covariates	With strata FE & covariates
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment group dummy	0.384***	0.301***	0.336***	0.288***	0.366***	0.303***
	(0.030)	(0.041)	(0.046)	(0.042)	(0.066)	(0.067)
F-test: strata FE		8.58		3.36		1.36
		[0.000]		[0.000]		[0.215]
F-test: covariates		21.13		11.78		4.30
		[0.000]		[0.000]		[0.000]
F-test: strata FE & covariates		89.86		10.33		4.49
		[0.000]		[0.000]		[0.000]
Mean of dependent variable	0.354	0.288	0.262	0.225	0.229	0.222
Comparison mean	0.068		0.073		0.021	
Observations	630	465	282	267	109	108

Notes: The dependent variable is a dummy for survey attritors. Comparison mean is the mean of each dependent variable in the comparison group. Strata fixed effects are constructed by training course and center dummies. Covariates include age, dummies for education level (including missing category), dummies for marital status (including missing category), dummies for types of disability (including missing category), years since demobilization, and a dummy for Forces Armées Rwandaises or armed group. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values of joint significance test (F test) are in brackets.

Table 2. Covariate balance

	Surveyed sample			Trimmed sample		
	Comparison group (1)	Treatment group (2)	p-value: (1)–(2) (3)	Comparison group (4)	Treatment group (5)	p-value: (4)–(5) (6)
Age	37.5	37.4	0.823	37.7	36.7	0.218
Education level						
Primary incomplete or below	0.100	0.105	0.871	0.123	0.074	0.242
Primary complete	0.040	0.214	0.000	0.053	0.170	0.009
Lower secondary incomplete	0.707	0.533	0.000	0.623	0.660	0.584
Lower secondary complete or above	0.113	0.062	0.090	0.149	0.021	0.001
Missing	0.040	0.086	0.055	0.053	0.074	0.526
Marital status						
Married	0.827	0.802	0.528	0.807	0.862	0.290
Missing	0.040	0.004	0.029	0.044	0.000	0.024
Type of disability						
Physical	0.533	0.568	0.498	0.561	0.596	0.619
Amputation	0.073	0.148	0.016	0.053	0.213	0.001
Blind	0.067	0.101	0.215	0.088	0.096	0.843
Other	0.227	0.136	0.026	0.219	0.064	0.001
Missing	0.100	0.047	0.057	0.079	0.032	0.134
Ex-army group						
Rwanda Defence Forces	0.753	0.747	0.888	0.763	0.617	0.024
Forces Armées Rwandaises	0.227	0.175	0.217	0.219	0.319	0.109
Armed group	0.020	0.078	0.005	0.018	0.064	0.102
Years since demobilization	6.852	7.031	0.542	6.965	6.543	0.225
Maximum number of observations	150	257		114	94	

Notes: Columns (3) and (6) report p-values of the test for the difference in means between the treatment and comparison groups based on robust standard errors.

Table 3. Treatment effects

	Economic empowerment			Social reintegration		
	Employed	Income (1000 RWF)	Income (log)	Treated preferably by family	Good relationships with neighbors	Treated the same/preferably in community
	(1)	(2)	(3)	(4)	(5)	(6)
Comparison mean	0.185	7.818	8.482	0.235	0.751	0.171
Panel A: Treatment effect	0.630***	12.818**	0.926**	0.344**	0.082	0.303***
(with no covariates)	(0.119)	(6.159)	(0.383)	(0.136)	(0.113)	(0.104)
	[0.001]	[0.013]	[0.022]	[0.001]	[0.426]	[0.054]
	{0.000}	{0.033}	{0.031}	{0.023}	{0.467}	{0.011}
Panel B: Treatment effect	0.641***	10.590	0.902***	0.391***	0.075	0.327***
(with strata FE & covariates)	(0.081)	(6.893)	(0.317)	(0.111)	(0.110)	(0.109)
	[0.002]	[0.032]	[0.029]	[0.003]	[0.503]	[0.034]
	{0.000}	{0.110}	{0.018}	{0.002}	{0.488}	{0.008}
Number of observations	80	62	62	63	74	73
Number of observations in trimmed sample (before matching)	203	141	141	162	194	187

Notes: The sample is the matched sample. Comparison mean is the mean of each outcome in the comparison group. Strata fixed effects are constructed by training course and center dummies. Covariates include age, dummies for education level (including missing category), dummies for marital status (including missing category), dummies for types of disability (including missing category), years since demobilization, and a dummy for Forces Armées Rwandaises or armed group. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values for wild-cluster bootstrap two-way cluster-robust standard errors (1,000 bootstrap replications) are in brackets. P-values for multiple hypothesis testing for each outcome category using robust standard errors (1,000 bootstrap replications) are in braces.

Table 4. Coefficient stability test

	Economic empowerment		Social reintegration	
	Employed	Income (log)	Treated preferably by family	Treated the same/preferably in community
	(1)	(2)	(3)	(4)
$\dot{\beta}$, Treatment effect (with no covariates)	0.630	0.926	0.344	0.303
R-squared ($\dot{R}$)	0.396	0.162	0.113	0.104
Observations	80	62	63	73
$\tilde{\beta}$, Treatment effect (with strata FE & covariates)	0.641	0.902	0.391	0.327
95% confidence interval	[0.480, 0.803]	[0.262, 1.541]	[0.168, 0.615]	[0.108, 0.546]
R-squared ($\tilde{R}$)	0.669	0.446	0.474	0.410
Observations	80	62	63	73
β^*, Bias-adjusted treatment effect				
$R_{max} = 1.3\tilde{R},\ \delta = 1$	0.649	0.890	0.412	0.337
$R_{max} = 1.3\tilde{R},\ \delta = -1$	0.633	0.914	0.372	0.317
$R_{max} = 2\tilde{R},\ \delta = 1$	0.654	0.859	0.465	0.362
$R_{max} = 2\tilde{R},\ \delta = -1$	0.628	0.940	0.329	0.294

Note: The sample is the matched sample. The 95% confidence intervals are obtained by robust standard errors.

Table 5. Lee bounds

	Economic empowerment		Social reintegration	
	Employed	Income (log)	Treated preferably by family	Treated the same/preferably in community
	(1)	(2)	(3)	(4)
Lower bound	0.614***	0.651**	0.301**	0.244*
	(0.095)	(0.291)	(0.126)	(0.136)
Upper bound	0.658***	1.018***	0.455***	0.368***
	(0.086)	(0.241)	(0.138)	(0.126)
q_T, Rate of obs. of treatment G	0.655	0.636	0.686	0.655
q_C, Rate of obs. of comparison G	1.000	1.000	1.000	1.000
q, Trimming proportion	0.345	0.364	0.314	0.345
Number of observations before matching	277	215	236	261
Number of observations after matching	100	81	76	93
Number of observations	80	61	60	73

Notes: The number of observations before matching is for surveyed trainees in the trimmed sample plus survey attritors within the same strata as those of the surveyed trainees; the number of observations after matching is for the matched sample including matched attritors used for computing the trimming proportions; and the number of observations is for the matched sample used for obtaining Lee bounds. Bootstrapped standard errors (500 bootstrap replications) are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Appendix

Table A1. Trainees by strata

	Enrolled trainees		Surveyed trainees		Matched trainees	
	Comparison group	Treatment group	Comparison group	Treatment group	Comparison group	Treatment group
ATC	19	42	18	26	6	6
Brick/block laying & concreting	5	10	5	5	3	3
Electronic	5	9	4	5	2	2
Hotel	1	3	1	2	0	0
Plumbing & pipe fitting	1	8	1	4	1	1
Welding	7	12	7	10	0	0
CFJ Kibali	27	30	24	22	11	4
Cooking	10	10	8	7	2	1
Welding	17	20	16	15	9	3
CFJ Nyanza	34	45	29	22	16	16
Brick/block laying & concreting	14	15	12	8	8	7
Plumbing & pipe fitting	10	15	8	6	2	3
Welding	10	15	9	8	6	6
CFJ Rwabuye	43	42	43	24	10	12
Brick/block laying & concreting	14	10	14	6	4	5
Carpentry & joinery	15	16	15	11	1	4
Tailoring	14	16	14	7	5	3
Total	123	159	114	94	43	38

Note: The table reports the number of trainees.

Table A2. Missingness

	Missing in income		Missing in "Treated preferably by family"	
	No covariates	With strata FE & covariates	No covariates	With strata FE & covariates
	(1)	(2)	(3)	(4)
Panel A: Surveyed sample				
Treatment group dummy	-0.263***	-0.339***	-0.362***	-0.376***
	(0.048)	(0.060)	(0.040)	(0.051)
F-test: strata FE		9.20		1.92
		[0.000]		[0.011]
F-test: covariates		3.31		1.61
		[0.001]		[0.081]
F-test: strata FE & covariates		68.86		1.75
		[0.000]		[0.008]
Mean of dependent variable	0.287	0.285	0.145	0.144
Comparison mean	0.453		0.373	
Observations	407	328	407	328
Panel B: Trimmed sample				
Treatment group dummy	-0.335***	-0.337***	-0.384***	-0.350***
	(0.059)	(0.063)	(0.047)	(0.056)
F-test: strata FE		2.89		2.26
		[0.001]		[0.011]
F-test: covariates		3.05		1.25
		[0.000]		[0.246]
F-test: strata FE & covariates		4.18		2.13
		[0.000]		[0.002]
Mean of dependent variable	0.322	0.319	0.221	0.217
Comparison mean	0.474		0.395	
Observations	208	207	208	207
Panel C: Matched sample				
Treatment group dummy	-0.343***	-0.306***	-0.419***	-0.408***
	(0.084)	(0.087)	(0.076)	(0.082)
F-test: strata FE		1.54		2.27
		[0.148]		[0.025]
F-test: covariates		2.76		1.06
		[0.009]		[0.408]
F-test: strata FE & covariates		3.68		2.06
		[0.000]		[0.018]
Mean of dependent variable	0.235	0.235	0.222	0.222
Comparison mean	0.395		0.419	
Observations	81	81	81	81

Notes: The dependent variable is a dummy for missing values. The comparison mean is the mean of each dependent variable in the comparison group. Strata fixed effects are constructed by training course and center dummies. Covariates include age, dummies for education level (including missing category), dummies for marital status (including missing category), dummies for types of disability (including missing category), years since demobilization, and a dummy for Forces Armées Rwandaises or armed group. Robust standard errors are in parentheses. ***p<0.01, **p<0.05, *p<0.1. P-values of joint significance test (F-test) are in brackets.

Table A3. Treatment effects, trimmed sample

	Economic empowerment			Social reintegration		
	Employed	Income (1000 RWF)	Income (log)	Treated preferably by family	Good relationships with neighbors	Treated the same/preferably in community
	(1)	(2)	(3)	(4)	(5)	(6)
Comparison mean	0.110	9.747	8.711	0.203	0.641	0.223
Panel A: Treatment effect	0.518***	6.949**	0.597***	0.324***	0.161**	0.282***
(with no covariates)	(0.058)	(3.217)	(0.168)	(0.071)	(0.063)	(0.068)
	[0.000]	[0.055]	[0.004]	[0.000]	[0.075]	[0.037]
	{0.000}	{0.031}	{0.003}	{0.000}	{0.017}	{0.000}
Panel B: Treatment effect	0.550***	9.866**	0.718***	0.315***	0.143*	0.308***
(with strata FE & covariates)	(0.065)	(3.844)	(0.219)	(0.084)	(0.078)	(0.083)
	[0.000]	[0.003]	[0.001]	[0.001]	[0.111]	[0.052]
	{0.000}	{0.019}	{0.007}	{0.002}	{0.059}	{0.002}
Number of observations	203	141	141	162	194	187

Notes: Comparison mean is the mean of each outcome in the comparison group. Strata fixed effects are constructed by training course and center dummies. Covariates include age, dummies for education level (including missing category), dummies for marital status (including missing category), dummies for types of disability (including missing category), years since demobilization, and a dummy for Forces Armées Rwandaises or armed group. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. P-values for wild-cluster bootstrap two-way cluster-robust standard errors (1,000 bootstrap replications) are in brackets. P-values for the multiple hypothesis testing for each outcome category using robust standard errors (1,000 bootstrap replications) are in braces.